**Recent advances in Complexity CIS 6930/CIS 4930** October 22, 2002

# Lecture 17

Lecturer: Dr. Meera Sitharam Scribe: Hongyu Guo

## 1 Erwin's talk on Natural Proofs (Part I)

Our goal is to show that there is no natural proof for the statement $P/poly \neq NP$ unless an unlikely cryptographical assumption holds.

First we give some notations and definitions. We denote by $F_n$ the set of Boolean functions in $n$ variables. $f_n$ is a randomly chosen function from $F_n$. Formally, by a combinatorial property of Boolean functions we mean a set of Boolean functions $\{C_n \subseteq F_n \mid n \in \omega\}$. Thus, a Boolean function $f_n$ possesses property $C_n$ $iff$ $f_n \in C_n$. (Alternatively, we will sometimes find it convenient to use function notation: $C_n(f_n) = 1$ if $f_n \in C_n$, and $C_n(f_n) = 0$ if $f_n \notin C_n$.)

**Definition 1** *A combinatorial property $C_n$ is **natural** if it contains a subset $C_n^*$ with the following two conditions:*

- ***Constructivity:** The predicate $f_n \in^? C_n^*$ is in P. Thus, $C_n^*$ is computable in time which is polynomial in the truth table of $f_n$;*
- ***Largeness:** $|C_n^*| \geq 2^{-O(n)} \cdot |F_n|$.*

**Definition 2** *A combinatorial property $C_n$ is **useful against** $P/poly$ if it satisfies:*

- ***Usefulness:** The circuit size of any sequence of functions $f_1, f_2, ..., f_n, ...$, where $f_n \in C_n$, is super-polymonial, ie., for any constant $k$, for sufficiently large $n$, the circuit size of $f_n$ is greater than $n^k$.*

A proof that some function does not have polynomial-sized circuits is *natural against P/poly* if the proof contains, more or less explicity, the definition of a natural combinatorial property $C_n$ which is useful against $P/poly$.

It is easy and useful to extend the definition of natural proof to a more general, parameterized version.

**Definition 3** *Let $\Gamma$ and $\Lambda$ be complexity classes. Call a combinatorial property $C_n$ $\Gamma$ – natural with desity $\delta_n$ if it contains $C_n^* \subseteq C_n$ with the following two conditions:*

- ***Constructivity:** The predicate $f_n \in^? C_n^*$ is computable in $\Gamma$ ($C_n^*$ is a set of truth-tables with $2^n$ bits);*
- ***Largeness:** $|C_n^*| \geq \delta_n \cdot |F_n|$.*

**Definition 4** *A combinatorial property $C_n$ is **useful against** $\Lambda$ if it satisfies:*

- ***Usefulness:*** *For any sequence of functions $f_n$, where the event $f_n \in C_n$ happens infinitely often, $\{f_n\} \notin \Lambda$.*

A lower bound proof that some explicit function is not in $\Lambda$ is called $\Gamma$ - *natural against* $\Lambda$ *with desity* $\delta_n$ if it states a $\Gamma$-natural property $C_n$ which is useful against $\Lambda$ with desity $\delta_n$.

The "default" settings of our parameters is $\Gamma = P, \Lambda = P/poly$, and $\delta_n = 2^{-O(n)}$, as in the initial definition. The main result implies the negative statement that, under the pseudo-randomness assumption, no proof with these parameters can show that SAT does not have polynomial-sized circuits.

**Example 1** *$AC^0$ lower bounds for parity: $AC^0$-natural*

*$C_n$ is the property that there does not exist a restriction of the variables with the appropriate number of unassigned variables which forces $f_n$ to be a constant function. Hastad Lemma says that $C_n(f_n) = 1$ implies that $\{f_n\} \notin AC^0$. In other words, that $C_n$ is useful against $AC^0$. We show that $C_n$ is a natural property. In fact, we can choose $C_n^* = C_n$.*

*$C_n^*$ has constructivity. $C_n^*$ is in $AC^0$. Suppose $k$ is the number of unassigned variables. Given the truth table for $f_n$ as input, we compute $C_n^*(f_n)$ as follows. List all $\binom{n}{k}2^{n-k} = 2^{O(n)}$ restrictions of $n-k$ variables. For each one there is a circuit of depth 2 and size $2^{O(n)}$ which outpusts a 1 iff that restriction does not leave $f_n$ a constant function. Output the AND of all these circuits. The resulting circuit has depth 3 and is polynomial-sized in $2^n$.*

*A simple counting argument shows $C_n^*$ has the largeness condition. First we count the size of $\bar{C}_n^*$, the complement set of $C_n^*$. Let $p = n-k$. We first consider a fixed set of variables, say $x_1, x_2, ..., x_p$. For a fixed assignment of these variables, e.g., $x_1 = 0, x_2 = 0, ..., x_p = 0$, there are $2 \cdot 2^{(2^n - 2^{n-p})}$ functions which are set to constant. Now consider all possible assignment for these fixed set of variables, there are at most $2^p \cdot 2 \cdot 2^{(2^n - 2^{n-p})}$ functions are set to constant. Note there are redundant count here and that's why we say "at most". (Later we need better approximations to get better bound.) There are $\binom{n}{p}$ sets of $p$ variables out of $n$ variables. There are at most $\binom{n}{p}2^p \cdot 2 \cdot 2^{(2^n - 2^{n-p})}$ functions are set to constant by any assignments to $p$ variables.*

$$|\bar{C}_n^*| \leq \binom{n}{p} 2^p \cdot 2 \cdot 2^{(2^n - 2^{n-p})}$$

$$\leq n^p \cdot 2^{p+1} \cdot 2^{(2^n - 2^{n-p})}$$

$$= n^p 2^{2^n - 2^{n-p} + p + 1}$$

*The fraction*
$\frac{|C_n^*|}{|F_n|}$
$= 1 - \frac{\bar{C}_n^*}{|F_n|}$
$= 1 - \frac{n^p 2^{2^n - 2^{n-p} + p + 1}}{2^{2^n}}$
$= 1 - n^p 2^{-2^{n-p} + p + 1}$
$\geq \frac{1}{2^{O(n)}}$ *(for large $n$ and small $p$)*

**Exercise 1** *What do the functions with property $C_n$ in Example 1 look like?*

Natural proofs for lower bounds are almost self-defeating. The idea is that a natural proof that some function $f$ is not in $P/poly$ has an associated algorithm. But just as the proof must distinguish $f$ from a pseudo-random function in $P/poly$, the associated algorithm must be able to tell the difference between the two. Thus, the algorithm can be used to break a pseudo-random generator. This is self-defeating in the sense that a natural proof that hardness exists would have as an automatic by-product an algorithm to solve a "hard" problem.

For a pseudo-random generator $G_k : \{0,1\}^k \longrightarrow \{0,1\}^{2k}$ define its *hardness* $H(G_k)$ as the minimal $S$ for which there exists a circuit $C$ of size $\leq S$ such that

$$|\mathbf{P}[C(G_k(\mathbf{x}) = 1] - \mathbf{P}[C(\mathbf{y}) = 1]| \geq \frac{1}{S}$$

Here, $\mathbf{x}$ is taken at random from $\{0,1\}^k$, and $\mathbf{y}$ is taken at random from $\{0,1\}^{2k}$.

Intuitively, the hardness of a pseudo-random generator is the smallest circuit size which the pseudo-random generator is not able to fool.

**Theorem 1** *There is no lower bound proof which is P/poly-natural against P/poly, unless $H(G_k) \leq 2^{k^{O(1)}}$ for every pseudo-random generator $G_k : \{0,1\}^k \longrightarrow \{0,1\}^{2k}$ in P/poly.*